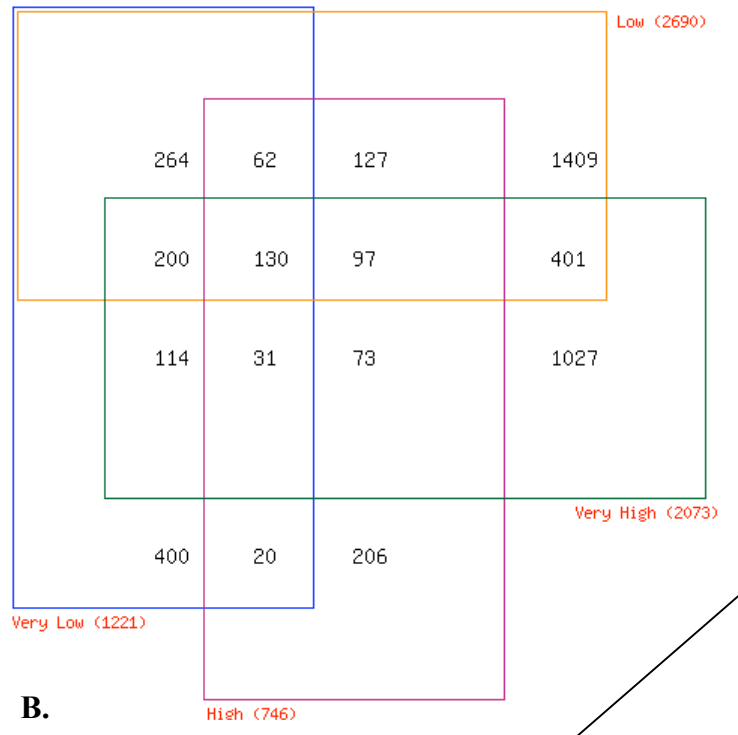
Supplemental Material

Global Gene Expression Profiling of a Population Exposed to a Range of Benzene Levels

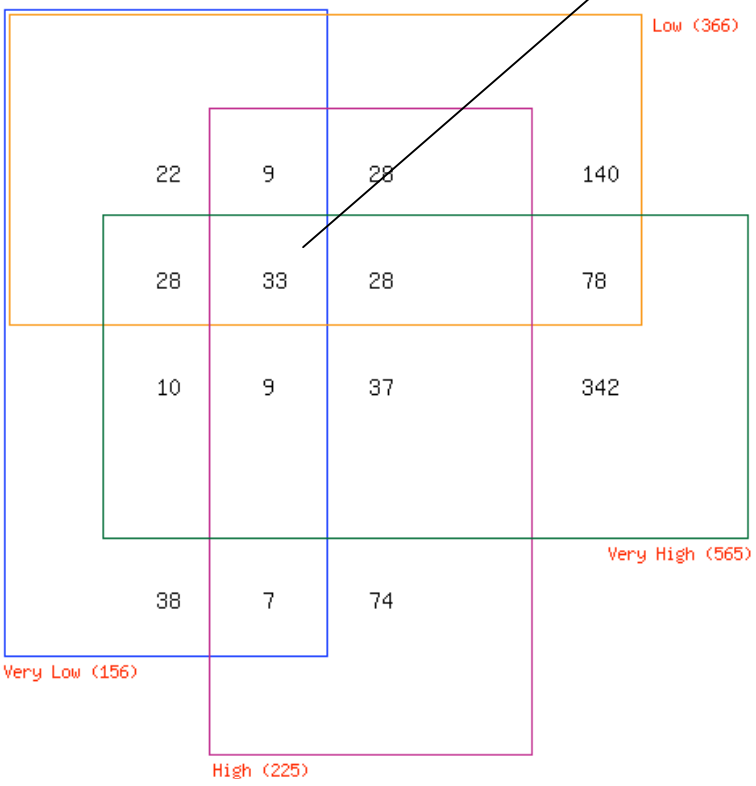
Cliona M. McHale, Luoping Zhang, Qing Lan, Roel Vermeulen Guilan Li, Alan E. Hubbard, Kristin E. Porter, Reuben Thomas, Christopher J. Portier, Min Shen, Stephen M. Rappaport, Songnian Yin, Martyn T. Smith, and Nathaniel Rothman.

Table of Contents

A.



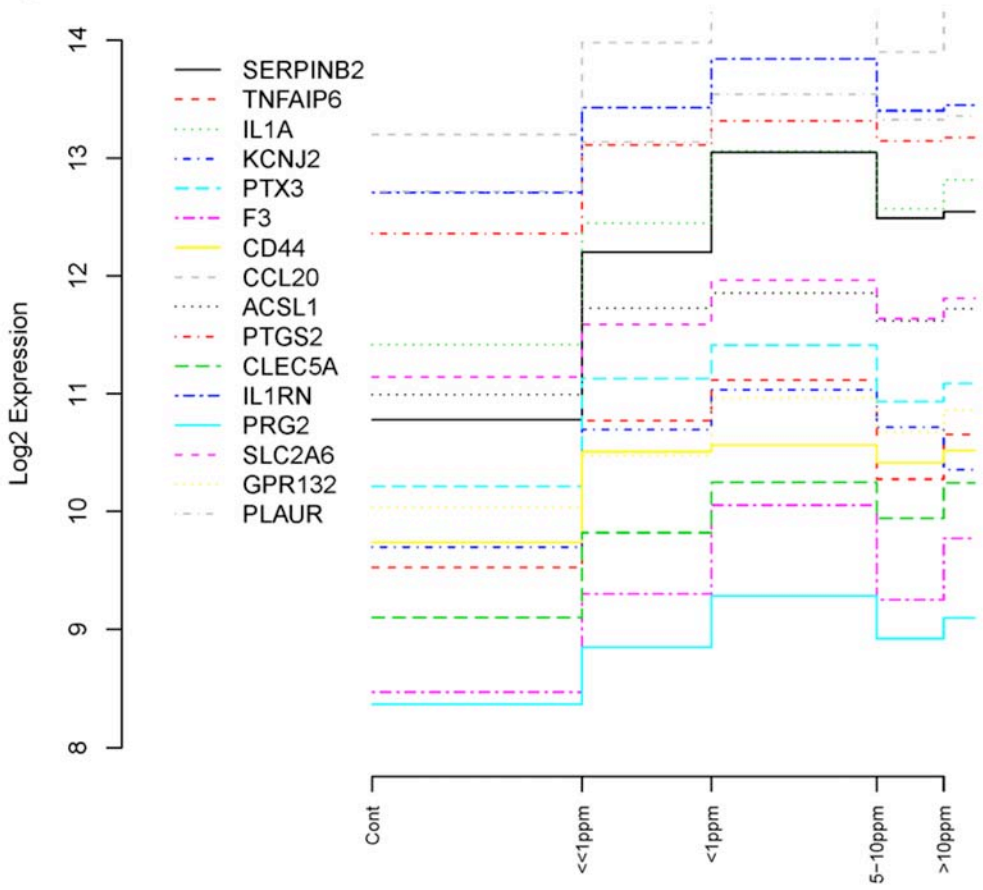
B.



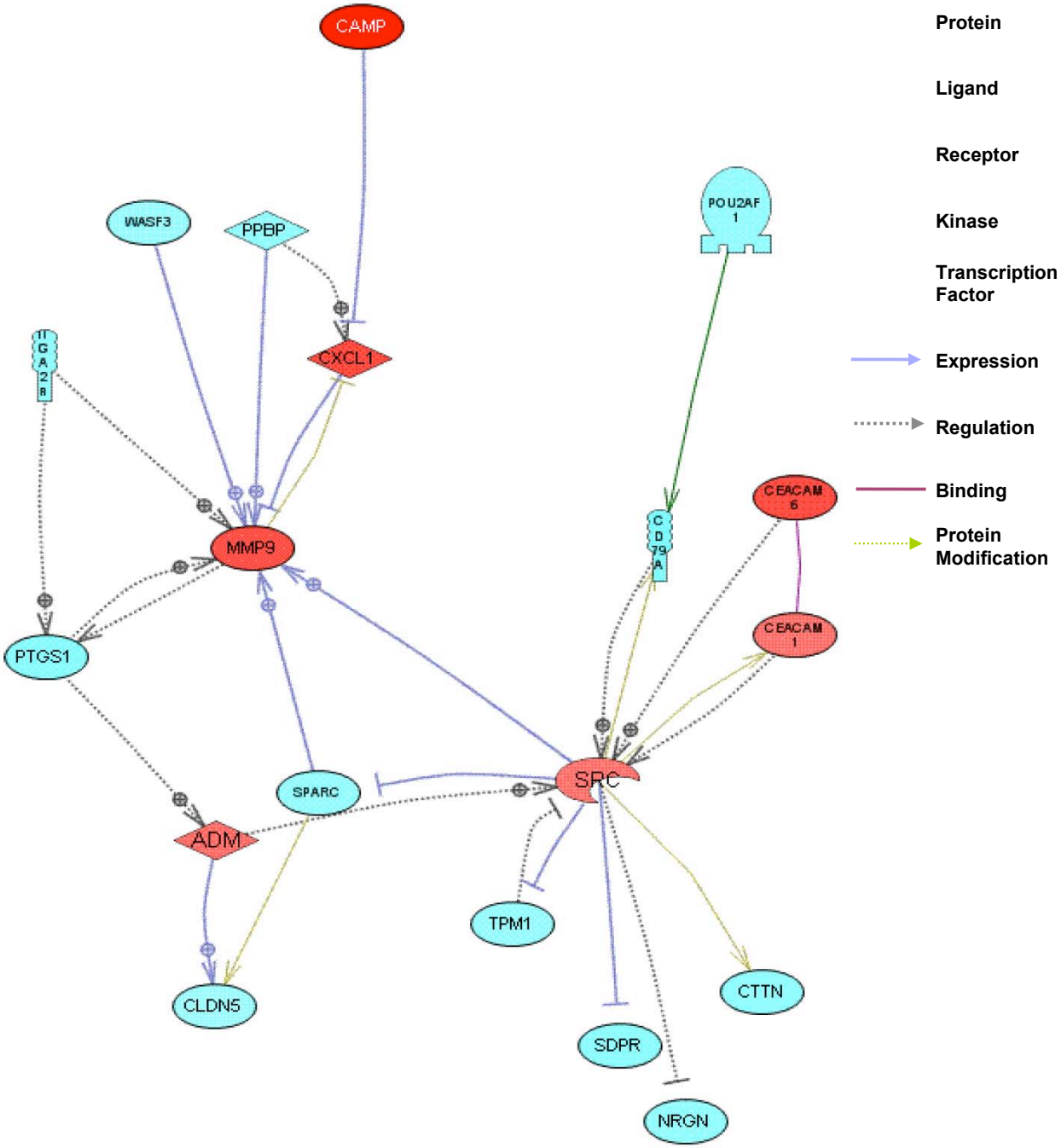
C.

Probe Id	Symbol
1030398	NFKB2
1260746	SAMSN1
1340092	IBRDC3
1410189	CLN8
1470682	SLC2A6
150632	NBN
1770152	KCNJ2
2120626	ATP2B1
2230131	CD44
2370524	PTGS2
2470100	CLEC5A
2850707	SERPINB9
2900286	PTX3
3310474	TNFAIP6
3710379	CCL20
4060674	SLC43A3
4230538	GPR132
430132	HSH2D
430619	PLAUR
4480202	IL1RN
4880717	LDLR
5090327	ARL5B
5130162	ACSL1
5550767	SERPINB2
5860333	MCOLN2
6020192	GCH1
6250743	F3
6380414	ADORA2A
6400537	GCH1
650709	AKNA
6590338	DCUN1D3
70048	IL1A
7320646	PRG2

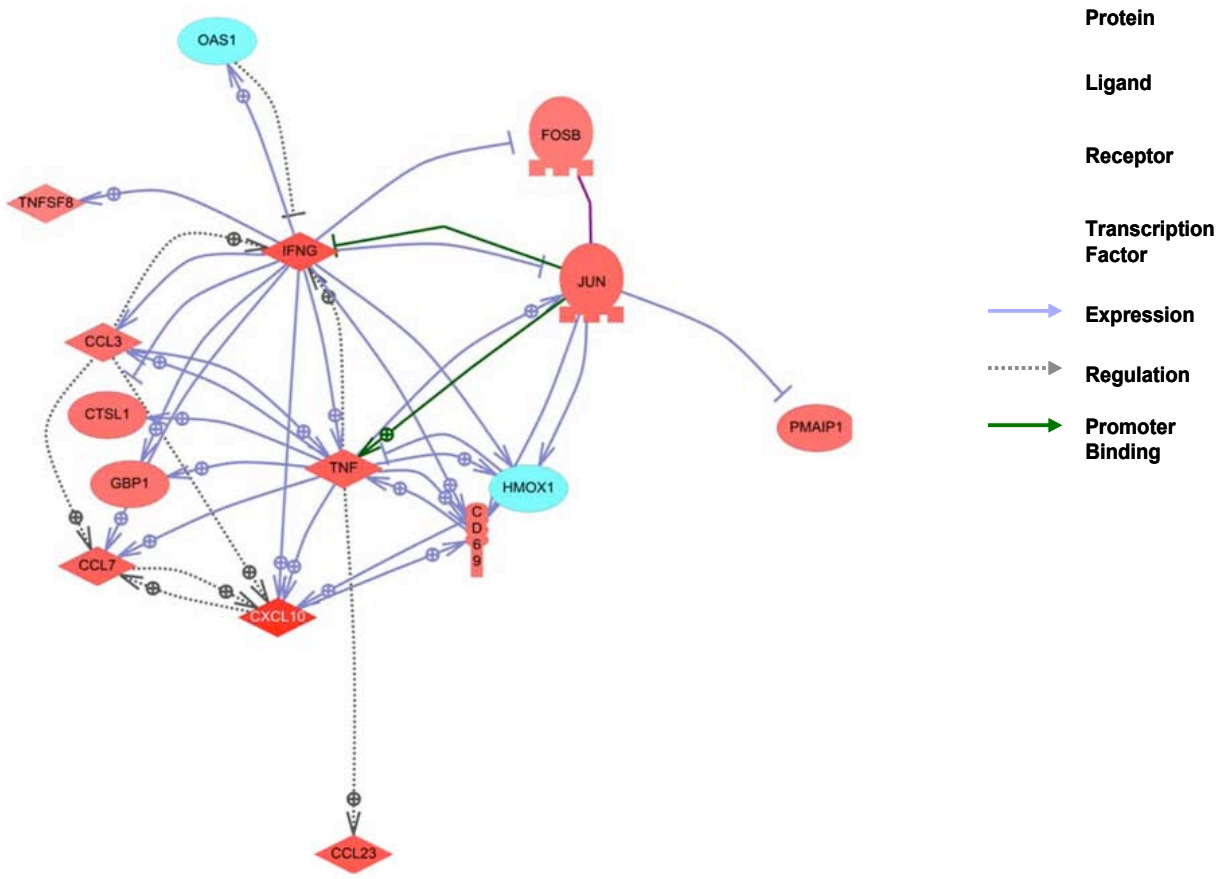
Supplemental Material, Figure 1. Overlap among genes significantly (FDR-adjusted *p*-values) altered (A) by any amount, (B) by ≥ 1.3-fold, at each exposure level, and (C) Genes commonly upregulated ≥ 1.3-fold at all 4 exposure levels.



Supplemental Material, Figure 2. Expression of potential biomarkers of benzene exposure plotted across exposure categories. The mean log2 expression value for each of the 16 potential biomarker genes, whose expression was up- or down-regulated ≥ 1.5-fold at 3 or 4 doses, are plotted for each exposure category. A distinct dose-response curve is common to all of these genes, with a peak in response at the <1 ppm category.



Supplemental Material, Figure 3. A network, involving 19 of the 78 potential biomarker genes unique to High-dose benzene exposure, centered round SRC and MMP9 was identified by Pathway Studio. Red and blue indicate up- and down-regulated gene expression.



Supplemental Material, Figure 4. A network, involving 15 of the 29 potential biomarker genes unique to Low-dose benzene exposure, centered round IFNG and TNF, was identified by Pathway Studio. Red and blue indicate up- and down-regulated gene expression.

Supplemental material, Table 1 is provided as a separate Excel file.

Randomization of samples and exposure levels across experimental variables. Samples were labeled samples in batches of 24 using Ambion's Illumina® RNA Amplification, kit http://ambion.com/techlib/prot/fm_IL1791.pdf, (Ambion, Austin, TX), and hybridized them to Illumina HumanRef-8 V2 BeadChips in batches of 32 (4 chips) following Illumina's protocol. All sample processing was performed in a blinded manner. Hyb, label and chip columns indicate the randomization of samples across these experimental stages. Dose Group shows the randomization by exposure category: Very High exposure (>10 ppm), High exposure (5-10 ppm), Low exposure (<1 ppm; average <1 ppm), Very Low exposure (<<1 ppm; average <1 ppm and most individual measurements below 1 ppm) and Control.

Supplemental material, Table 2 is provided as a separate Excel file.

Significant probes associated with benzene overall i.e. across one or more of four exposure categories, Very High exposure (>10 ppm), High exposure (5-10 ppm), Low exposure (<1 ppm; average <1 ppm), Very Low exposure (<<1 ppm; average <1 ppm and most individual measurements below 1 ppm) relative to Controls. ProbeID lists Illumina identifiers for the Illumina HumanRef-8_V2_11223162_B BeadChip. *P*-value indicates raw *p*-values from the mixed effects model; FDR (false discovery rate) indicates p-values adjusted with the Benjamin-Hochberg procedure (Benjamini and Hochberg 1995); Accessions and symbols are from GenBank (http://www.ncbi.nlm.nih.gov/genbank/).

Supplemental material, Table 3. Gene Ontology Processes Over-represented among the 3007 Genes Associated with Benzene Exposure Overall.

GO term	GO category	*p*-value*	No. genes
GO:0006955	immune response	3.78E-07	146 (661)**
GO:0006506	GPI anchor biosynthetic process	2.77E-05	13 (26)
GO:0045410	positive regulation of interleukin-6 biosynthetic process	7.46E-05	5 (5)
GO:0006917	induction of apoptosis	2.04E-04	56 (235)
GO:0006512	ubiquitin cycle	2.04E-04	111 (539)
GO:0006412	translation	2.70E-04	106 (456)
GO:0015986	ATP synthesis coupled proton transport	1.49E-03	11 (28)
GO:0006954	inflammatory response	1.57E-03	71 (338)
GO:0042113	B cell activation	1.72E-03	20 (68)

*p-value is from the Fisher's exact test in the elim method in TopGO (Alexa et al. 2006) (http://www.bioconductor.org/packages/bioc/html/topGO.html); ** No. significant genes in the GO category (total number of annotated genes in the GO category that were included on the chip).

Supplemental material, Table 4. Pathways Over-represented among the 3007 Genes Associated with Benzene Exposure Overall.

KEGG ID*	Pathway	*p*-value**
hsa04620	Toll-like receptor signaling pathway	<0.001
hsa04210	Apoptosis	<0.001
hsa05221	Acute myeloid leukemia	<0.001
hsa00190	Oxidative phosphorylation	<0.001
hsa04662	B cell receptor signaling pathway	<0.001
hsa04660	T cell receptor signaling pathway	0.001
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	0.002
hsa04060	Cytokine-cytokine receptor interaction	0.003
hsa00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	0.003
hsa05222	Small cell lung cancer	0.004

*KEGG IDs are from http://www.genome.jp/kegg/pathway.html; ***p*-values from the SEPEA_NT3 pathway enrichment analysis method (Thomas et al. 2009), *p*-values <0.005 correspond to a Bonferroni adjustment for multiple testing.

Supplemental material, Table 5 is provided as a separate Excel file.

Significant probes associated with Very Low Dose benzene exposure (average <1 ppm and most individual measurements below 1 ppm) vs control

ProbeID lists Illumina identifiers for the Illumina HumanRef-8_V2_11223162_B BeadChip. *P*-value indicates raw *p*-values from the mixed effects model; FDR (false discovery rate) indicates p-values adjusted with the Benjamin-Hochberg procedure (Benjamini and Hochberg 1995); Accessions and symbols are from GenBank (http://www.ncbi.nlm.nih.gov/genbank/).

Supplemental material, Table 6 is provided as a separate Excel file.

Significant probes associated with Low Dose benzene exposure (average <1 ppm) vs Control. ProbeID lists Illumina identifiers for the Illumina HumanRef-8_V2_11223162_B BeadChip. *P*-value indicates raw *p*-values from the mixed effects model; FDR (false discovery rate) indicates p-values adjusted with the Benjamin-Hochberg procedure (Benjamini and Hochberg 1995); Accessions and symbols are from GenBank (http://www.ncbi.nlm.nih.gov/genbank/).

Supplemental material, Table 7 is provided as a separate Excel file.

Significant probes associated with High Dose benzene exposure (5-10 ppm) vs Control. ProbeID lists Illumina identifiers for the Illumina HumanRef-8_V2_11223162_B BeadChip. *P*-value indicates raw *p*-values from the mixed effects model; FDR (false discovery rate) indicates p-values adjusted with the Benjamin-Hochberg procedure (Benjamini and Hochberg 1995); Accessions and symbols are from GenBank (http://www.ncbi.nlm.nih.gov/genbank/).

Supplemental material, Table 8 is provided as a separate Excel file.

Significant probes associated with Very High Dose benzene exposure (>10 ppm) vs control. ProbeID lists Illumina identifiers for the Illumina HumanRef-8_V2_11223162_B BeadChip. *P*-value indicates raw *p*-values from the mixed effects model; FDR (false discovery rate) indicates p-values adjusted with the Benjamin-Hochberg procedure (Benjamini and Hochberg 1995); Accessions and symbols are from GenBank (http://www.ncbi.nlm.nih.gov/genbank/).

Supplemental material, Table 9. Summary of GO Processes Overrepresented at each Benzene Exposure Category

			Very Low		Low		High		Very High	
			(n = 29)		(n = 30)		(n = 11)		(n = 13)	
GO Term	**GO Term**	**Total No. genes***	**No. genes**	***p*** **	**No. genes**	***p***	**No. genes**	***p***	**No. genes**	***p***
GO:0048193	Golgi vesicle transport	133	19	5.6E-04						
GO:0006886	intracellular protein transport	418	50	3.7E-05						
GO:0008089	anterograde axon cargo transport	5	3	2.2E-03						
GO:0008380	RNA splicing	223	26	1.5E-03						
GO:0006412	translation	456	64	2.0E-06	93	1.2E-03				
GO:0006512	ubiquitin cycle	480	48	7.5E-04	98	1.6E-05				
GO:0006511	ubiquitin-dependent protein catabolic process	174			44	2.3E-05				
GO:0045410	positive regulation of interleukin-6 biosynthetic process	7			5	7.6E-04				
GO:0030512	negative regulation of transforming growth factor beta signaling pathway	5			4	1.5E-03				
GO:0019221	cytokine and chemokine mediated signaling pathway	38			13	1.0E-03				
GO:0006917	induction of apoptosis	216	27	4.1E-04	49	1.6E-04	*19*	*1.5E-03*		
GO:0006955	immune response	653	*58*	*3.7E-03*	124	4.6E-05	54	4.9E-06	97	1.1E-04
GO:0015986	ATP synthesis coupled proton transport	40	11	2.2E-05	14	5.0E-04			11	1.8E-03
GO:0006915	apoptosis	804	80	5.6E-03	158	9.2E-04			107	2.7E-03
GO:0030301	cholesterol transport	8	5	4.4E-05	*4*	*1.5E-02*			*4*	*5.5E-03*
GO:0006954	inflammatory response	318			*60*	*4.6E-03*	34	2.8E-05		
GO:0042108	positive regulation of cytokine biosynthetic process	42					9	5.6E-04		
GO:0045408	regulation of interleukin-6 biosynthetic process	12					4	7.8E-04		
GO:0007179	transforming growth factor beta receptor signaling pathway	57					8	1.3E-03		

Supplemental material, Table 9. Summary of GO Processes Overrepresented at each Benzene Exposure Category (contd.)

			Very Low		Low		High		Very High	
			(n = 29)		(n = 30)		(n = 11)		(n = 13)	
GO.ID	**Term**	**Total No. genes***	**No. genes**	***p*******	**No. genes**	***p***	**No. genes**	***p***	**No. genes**	***p***
GO:0009967	positive regulation of signal transduction	198					17	1.4E-03		
GO:0006166	purine ribonucleoside salvage	6							4	1.4E-03
GO:0006334	nucleosome assembly	79							17	2.4E-03
GO:0009058	biosynthetic process	1032							185	9.3E-04
GO:0009611	response to wounding	433							63	2.7E-03
GO:0042787	protein ubiquitination during ubiquitin-dependent protein catabolic process	7							4	3.0E-03
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	7							21	2.9E-03

*Number of annotated genes included on the chip, ***p*-value is from the Fisher's exact test in the elim method in TopGO (Alexa et al. 2006) (http://www.bioconductor.org/packages/bioc/html/topGO.html); Italics are not key nodes (elim method) but are significant by the classic method in TopGO.

Supplemental material, Table 10. Pathways Impacted at each Benzene Exposure Category

KEGG ID*	KEGG Pathway Name	*p*-value**			
		Very Low (n = 29)	Low (n = 30)	High (n = 11)	Very High (n = 13)
hsa00230	Purine metabolism	0.011			
hsa04810	Regulation of actin cytoskeleton	0.033			
hsa05220	Chronic myeloid leukemia	0.034	0.033		
hsa05212	Pancreatic cancer	0.023	0.007		
hsa00190	Oxidative phosphorylation ***	<0.001	0.003	0.001	
hsa05222	Small cell lung cancer ***	0.004	0.002	0.027	
hsa04662	B cell receptor signaling pathway***	0.008	0.003	0.004	
hsa04910	Insulin signaling pathway	0.015	0.035	0.052	
hsa04920	Adipocytokine signaling pathway	0.034	0.002	0.019	
hsa04710	Circadian rhythm - mammal	0.04	0.045	0.004	
hsa03020	RNA polymerase	<0.001		0.048	
hsa04620	Toll-like receptor signaling pathway***	<0.001	0.002	0.001	0.004
hsa05120	Epithelial cell signaling in Helicobacter pylori infection ***	<0.001	0.003	0.006	0.011
hsa00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis ***	<0.001	0.041	<0.001	0.007
hsa04660	T cell receptor signaling pathway ***	0.005	0.002	0.005	0.018
hsa04210	Apoptosis ***	0.007	0.002	0.007	0.013
hsa04060	Cytokine-cytokine receptor interaction ***	0.036	0.011	0.030	0.004
hsa05221	Acute myeloid leukemia ***	0.037	0.002		0.045
hsa00071	Fatty acid metabolism	0.037		0.049	0.033
hsa03420	Nucleotide excision repair	0.001		0.008	0.005
hsa03320	PPAR signaling pathway		0.04		
hsa04350	TGF-beta signaling pathway		0.005		
hsa04370	VEGF signaling pathway		0.029		
hsa04640	Hematopoietic cell lineage		0.021		
hsa00770	Pantothenate and CoA biosynthesis		0.007		

Supplemental material, Table 10. Pathways Impacted at each Benzene Exposure Category (contd.)

KEGG ID*	KEGG Pathway Name	*p*-value**			
		Very Low (n = 29)	Low (n = 30)	High (n = 11)	Very High (n = 13)
hsa04650	Natural killer cell mediated cytotoxicity		0.007		
hsa04912	GnRH signaling pathway		0.011		
hsa04940	Type I diabetes mellitus		0.013		
hsa05211	Renal cell carcinoma		0.024	0.015	
hsa03060	Protein export		0.053	0.024	
hsa00532	Chondroitin sulfate biosynthesis			0.047	
hsa00100	Steroid biosynthesis			0.004	0.034
hsa04664	Fc epsilon RI signaling pathway		0.006		0.046
hsa04630	Jak-STAT signaling pathway		0.003		0.048
hsa04010	MAPK signaling pathway		0.009		0.023
hsa04520	Adherens junction				0.036
hsa00533	Keratan sulfate biosynthesis				0.036
hsa04930	Type II diabetes mellitus				0.041
hsa03030	DNA replication				0.031
hsa04720	Long-term potentiation				0.024
hsa02010	ABC transporters				<0.001
hsa00290	Valine, leucine and isoleucine biosynthesis				0.014

*KEGG pathway names are from http://www.genome.jp/kegg/pathway.html, ***p*-values from the SEPEA_NT3 pathway enrichment analysis method (Thomas et al. 2009), *p*-values <0.005 correspond to a Bonferroni adjustment for multiple testing; ***significant in overall pathway analysis (Supplemental Material, Table 3).

Supplemental Material, Table 11. Functions of the Potential Biomarkers of Benzene Exposure

Symbol	Associated GO terms	Description	Function	Cellular component
SERPINB2	anti-apoptosis, wound healing	member of family of serine protease inhibitors, a major product of activated monocytes and macrophages	substantially induced during most inflammatory processes with proposed roles in the regulation of inflammation and wound healing	cytoplasm, extracellular region, extracellular space
TNFAIP6	cell adhesion, cell-cell signaling, inflammatory response, signal transduction	a member of the hyaluronan-binding protein family, induced by tumor necrosis factor alpha and interleukin-1	involved in extracellular matrix stability and cell migration	
IL1A	anti-apoptosis, apoptosis, cytokine-mediated signaling, fever, immune response, inflammatory response, negative regulation of cell proliferation, positive regulation of angiogenesis, positive regulation of cell division, positive regulation of cytokine secretion, positive regulation of mitosis, positive regulation of VEGF growth factro production	pleiotropic cytokine involved in various immune responses, inflammatory processes, and hematopoiesis	produced by monocytes and macrophages as a proprotein, is proteolytically processed and released in response to cell injury, and thus induces apoptosis	cytoplasm, extracellular component, extracellular space
KCNJ2	ion transport, potassium ion transport	integral membrane protein and inward-rectifier type potassium channel.	important regulators of resting membrane potential and cell excitability. responsible with other Kir2.x channels for the inward rectifier current IK1 in the human heart ventricle.	integral to plasma membrane

Supplemental Material, Table 11. Functions of Potential Biomarkers of Benzene Exposure (contd.)

Symbol	Associated GO terms	Description	Function	Cellular component
PTX3	inflammatory response, positive regulation of nitric oxide biosynthetic process, positive regulation of phagocytosis	long pentraxin (short pentraxin is CRP), produced by a variety of cells and tissues, most notably dendritic cells and macrophages, in response to Toll-like receptor (TLR) engagement and inflammatory cytokines	key components of the humoral arm of innate immunity and play a role in controlling inflammation. role in complement activation, pathogen recognition and apoptotic cell clearance	extracellular region
F3	activation of blood coagulation via clotting cascade, activation of caspase activity, activation of plasma proteins involved in acute inflammatory response, anti-apoptosis, positive regulation of angiogenesis, positive regulation of endothelial cell migration, positive regulation of PDGF receptor signaling pathway, positive regulation of positive chemotaxis, positive regulation of protein kinase B signaling cascade	coagulation factor III, a cell surface glycoprotein	initiates the blood coagulation cascades	extracellular matrix, extracellular space, integral to external side of plasma membrane,
CD44	blood vessel maturation, cell migration, cell-cell adhesion, cell-matrix adhesion, healing during inflammatory response, positive regulation of neurtophil apoptosis, regulation of cell growth	receptor for hyaluronic acid (HA), can also interact with other ligands, such as osteopontin, collagens, and matrix metalloproteinases (MMPs).	lymphocyte activation, recirculation and homing, hematopoiesis, and tumor metastasis. cell-surface glycoprotein involved in cell-cell interactions, cell adhesion and migration.	plasma membrane, cytoplasm, extracellular matrix, focal adhesion, Golgi apparatus, nucleus

Supplemental Material, Table 11. Functions of Potential Biomarkers of Benzene Exposure (contd.)

Symbol	Associated GO terms	Description	Function	Cellular component
CCL20	cell-cell signaling, chemokinesis, chemotaxis, immune response, inflammatory response, signal transduction	a small cytokine belonging to the CC chemokine family, induced by microbial factors such as lipopolysaccharide (LPS), and inflammatory cytokines such as tumor necrosis factor and interferon-γ, and down-regulated by IL-10 .	strongly chemotactic for lymphocytes and weakly attracts neutrophils. Elicits its effects on its target cells by binding and activating the chemokine receptor CCR6	extracellular region, extracellular space
ACSL1	fatty acid metabolic process, regulation of fatty acid oxidation	isozyme of the long-chain fatty-acid-coenzyme A ligase family	convert free long-chain fatty acids into fatty acyl-CoA esters, plays a key role in lipid biosynthesis and fatty acid degradation	cytosol, ER, integral to membrane, mitochondrial outer membrane, peroxisomal membrane,
PTGS2	cyclooxygenase pathway, fatty acid biosynthesis, oxidoreductive process, positive regulation of apoptosis, regulation of blood pressure, regulation of cell cycle, regulation of inflammatory response, response to oxidative stress	the key enzyme in prostaglandin biosynthesis, and acts both as a dioxygenase and as a peroxidase. encodes the inducible isozyme	responsible for the prostanoid biosynthesis involved in inflammation and mitogenesis	cytoplasm ER lumen, extrinsic to membrane, microsome, nucleus
CLEC5A	immune response, cellular defense response, signal transduction	encodes a member of the C-type lectin/C-type lectin-like domain (CTL/CTLD) superfamily	members have diverse functions, such as cell adhesion, cell-cell signalling, glycoprotein turnover, and roles in inflammation and immune response	integral to plasma membrane

Supplemental Material, Table 11. Functions of Potential Biomarkers of Benzene Exposure (contd.)

Symbol	Associated GO terms	Description	Function	Cellular component
IL1RN	immune response, inflammatory response	member of the interleukin 1 cytokine family	inhibits the activities of interleukin 1, alpha (IL1A) and interleukin 1, beta (IL1B), and modulates a variety of interleukin 1 related immune and inflammatory responses	cytoplasm, extracellular space
PRG2	immune response, defense response to bacterium	the predominant constituent of the crystalline core of the eosinophil granule	may be involved in antiparasitic defense mechanisms as a cytotoxin and helminthotoxin, and in immune hypersensitivity reactions	extracellular region
SLC2A6	carbohydrate transport	Hexose transport into mammalian cells is catalyzed by a family of membrane proteins, including SLC2A6		plasma membrane
GPR132		member of the G-protein couple receptor (GPCR) superfamily. High-affinity receptor for lysophosphatidylcholine (LPC), a major phospholipid component of oxidized low density lipoprotein.	This protein may react to LPC levels at sites of inflammation to limit the expansion of tissue-infiltrating cells.	

Supplemental Material, Table 11. Functions of Potential Biomarkers of Benzene Exposure (contd.)

Symbol	Associated GO terms	Description	Function	Cellular component
PLAUR	attachment of GPI anchor to protein, blood coagulation, chemotaxis	encodes the receptor for urokinase plasminogen activator and, role in localizing and promoting plasmin formation, likely influences many normal and pathological processes related to cell-surface plasminogen activation and localized degradation of the extracellular matrix.	lacks transmembrane or cytoplasmic domains, may be anchored to the plasma membrane by a glycosyl-phosphatidylinositol (GPI) moiety. A soluble protein is also produced in some cell types.	anchored to membrane, plasma membrane

Information derived from EntrezGene (http://www.ncbi.nlm.nih.gov/gene)

References

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22(13):1600-1607.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B(57):289-300.

Thomas R, Gohlke JM, Stopper GF, Parham FM, Portier CJ. 2009. Choosing the right path: enhancement of biologically relevant sets of genes or proteins using pathway structure. Genome Biol 10(4):R44.